

## 構造化文書処理装置

特開平7-56920

(19)日本国特許庁 (JP)

## (12) 公開特許公報 (A)

(11)特許出願公開番号

特開平7-56920

(43)公開日 平成7年(1995)3月3日

(51)Int.Cl.<sup>6</sup>  
G 0 6 F 17/27

識別記号

庁内整理番号

F I

技術表示箇所

7315-5L

G 0 6 F 15/ 20

5 5 0 E

審査請求 未請求 請求項の数4 FD (全16頁)

(21)出願番号 特願平5-218989

(22)出願日 平成5年(1993)8月11日

(71)出願人 000005496

富士ゼロックス株式会社

東京都港区赤坂三丁目3番5号

(72)発明者 陌間 端

神奈川県横浜市保土ヶ谷区神戸町134番地

横浜ビジネスパークイーストタワー 富士ゼロックス株式会社内

(72)発明者 畑中 正明

神奈川県川崎市高津区坂戸3丁目2番1号

K S P R &amp; D ビジネスパークビル

富士ゼロックス株式会社内

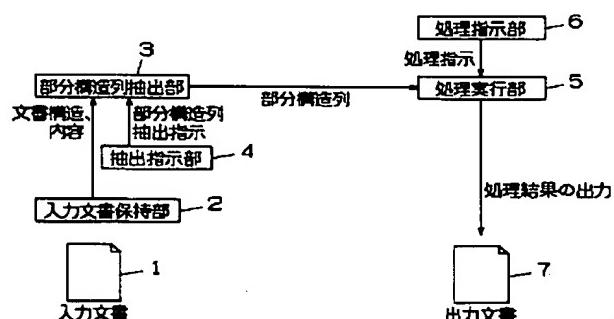
(74)代理人 弁理士 石井 康夫 (外1名)

(54)【発明の名称】 構造化文書処理装置

## (57)【要約】

【目的】 ユーザに繁雑な作業を強いることなく、構造化文書の複数の文書部品に対して一括した文書処理を行なうことのできる構造化文書処理装置を提供する。

【構成】 入力文書保持部2は、構造化文書である入力文書1を保持し、部分構造列抽出部3に対して、入力文書1の構造および内容を引き渡す。部分構造列抽出部3は、抽出指示部4からの抽出指示に基づいて、入力文書1の構造から、部分構造を抜き出し、部分構造列を生成する。処理実行部5は、部分構造列抽出部3において生成された部分構造列、及び、処理指示部6の処理指示に基づいて、構造化文書処理を実行し、その結果を出力文書7として出力する。



## 【特許請求の範囲】

【請求項1】 構造化文書に対して処理を行なう構造化文書処理装置において、入力文書から所定の条件に従って部分構造を抜き出しこれに基づき部分構造列を生成する部分構造抽出手段と、該部分構造列抽出手段により生成された前記部分構造列を用いて指示された処理を行ない出力文書を作成する処理手段を有することを特徴とする構造化文書処理装置。

【請求項2】 出力文書のひな型となるテンプレート文書を保持するテンプレート文書保持手段を有し、前記処理手段は、前記部分構造列抽出手段により生成された前記部分構造列を用い前記テンプレート文書保持手段に保持されている前記テンプレート文書に基づき指示された処理を行ない出力文書を作成することを特徴とする請求項1に記載の構造化文書処理装置。

【請求項3】 処理の指示あるいは前記テンプレート文書保持手段に保持されている前記テンプレート文書の情報に従い、前記部分構造列抽出手段により生成された前記部分構造列に対して処理を行なう部分構造列処理手段を有することを特徴とする請求項2に記載の構造化文書処理装置。

【請求項4】 前記テンプレート文書保持手段に保持されている前記テンプレート文書に対して、処理の指示あるいは前記部分構造列抽出手段により生成された前記部分構造列に基づいた処理を行なうテンプレート文書処理手段を有することを特徴とする請求項2または3に記載の構造化文書処理装置。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は、構造化文書の処理を行なう構造化文書処理装置に関するものである。

## 【0002】

【従来の技術】業務上の文書を作成する場合、そのひな型となる帳票を用意しておき、そこに必要な内容を記入するようにすることは一般に行なわれている。記入すべき内容が文書の作成者が考案しなければならない内容ではなく、日付けや、その他何らかの条件で決定されるものであったり、他の文書にある内容をそのまま複写したりすることもある。文書を文字列として扱う従来の文書処理装置では、このような文書の作成者が内容を考えるものではなく、ある程度自動的に処理可能な編集操作まで文書の作成者が行なわなければならず、繁雑である。

【0003】一方、文書を内容の単位で階層化した構造化文書が注目されている。構造化文書の代表的なものとしては、例えば、“ISO／IS 8613, Information Processing—Text and Office System—Office Document Architecture (ODA) and Interchange Format (1988)”や、Furuta. R., “Concept

s and Model for Structured Documents”, Cambridge University Press, pp. 7-38 (1988) に記載されたものなどが知られている。構造化文書を用いると、論理構造のひな型をあらかじめ用意しておき、ひな型の文書中の論理的な単位（文書部品）を挿入・置換する文書処理を行なって、文書を作成することが可能となる。これにより、上述のような、文書の作成者が内容を考えなくても良い部分について、ある程度の自動的な処理が可能となる。

【0004】このような文書処理を行なう場合、文書中の複数の論理部品を簡便に扱うことにより、利便性が向上する。例えば、ある文書中の図表を別の文書に複製して再利用しようとする場合を考える。そのような処理を、再利用する図表の1つ1つについて指示を与える必要がある処理装置を用いて実現するとすれば、ユーザの手間は非常に大きいものとなる。そのような装置では、従来のエディター上で、コピー、ペーストを繰り返すことと比べて、利便性が向上したとはいえない。

【0005】例えば、特願平4-207620号には、構造化文書中の複数の文書部品を同時に操作する方法が示されている。この方法では、複数の文書中の複数の文書部品を一括して取り扱うために、共通の属性値を用いる。共通の属性値を各文書部品にあらかじめ設定しておくことにより、ユーザの意図する処理動作を共通の属性値が設定されている各文書部品に対して一括して行なうことが可能となった。この方法を用いることにより、ユーザが、処理対象の文書部品を特定する属性値と、その属性値に対応する処理動作との組をシステムに与えることにより、簡便に構造化文書に対する処理を行なうことができる。このとき、処理対象となる文書部品を特定する属性値を切り替えることにより、どのような処理を行なうかを切り替えることができる。

【0006】しかしながら、この方法では、文書の作成者が属性値を管理しなければならない。大量の文書部品を処理の対象とする場合、文書の作成者は、処理の種類ごとに別の属性値を各文書部品に付加する必要があり、はなはだ煩雑である。また、文書部品を新たに作成した際に、どのような処理に用いるのかを考慮して、処理に応じた属性値を付加しなければならない。

【0007】この従来の方法による処理を、例をもとに説明する。入力文書から表を抜き出し、ひな型の文書であるテンプレート文書の各章の末尾に挿入する処理を例にする。図35は、従来の表の挿入処理を説明するための第1の例の説明図、図36は、図35における入力文書の論理構造の説明図、図37は、図35におけるテンプレート文書の論理構造の説明図である。図35において、入力文書は2つの表からなる。入力文書の論理構造を図36に示す。また、テンプレート文書は2つの章からなる。テンプレート文書の論理構造を図37に示す。

す。図35に示したテンプレート文書および図37に示したその構造のうち、点線で示した文書部品は、テンプレート文書が用意された段階では存在せず、文書処理の結果生じる部分を示している。以下の図中においても、点線を用いて表現する文書部品は、同様の意味に用いる。図35に示したテンプレート文書の各章の末尾に、点線で示したように、入力文書からの表をそれぞれ挿入する処理を実現することを考える。

【0008】上述の特願平4-207620号に記載されている方法によれば、処理の種類ごとに各文書部品に統一した属性値を付与する。この例では、入力文書の文書部品である「表1」および「表2」に、例えば、「挿入部品」という属性値をそれぞれ付与し、かつ、テンプレート文書の第1章の「本文」および第2章の「本文」に、例えば、「末子に挿入」という属性値をそれぞれ付与する。付与された属性値に基づいて統一した処理が実行され、この場合、表の挿入処理が実現する。

【0009】しかしながら、入力文書およびテンプレート文書が変更された場合に問題が生じる。図38は、従来の表の挿入処理を説明するための第2の例の説明図、図39は、図38における入力文書の論理構造の説明図、図40は、図38におけるテンプレート文書の論理構造の説明図である。図38に示すように、ユーザによる編集によって、入力文書に「表3」が加えられて、表の数が3つに増え、また、テンプレート文書に「第3章」が追加されたとする。このとき、処理の種類ごとに各文書部品に統一した属性値を付与するために、図39の入力文書の論理構造中の「表3」、および、図40のテンプレート文書の論理構造中の第3章の「本文」に属性値を付与する必要がある。どのような属性値が必要かを知るためにには、以前に属性値を付与されている文書部品、例えば、「表1」や、第1章の「本文」等を参照しなければならない。参照した結果、「挿入部品」、「末子に挿入」という属性値を付与しなければいけないことがわかるので、それぞれの文書部品にこれらの属性値を新たに付与することになる。このように、処理に用いる文書の変更の度に、属性値の付与という、はなはだ煩雑な作業を行なわなければならない。

【0010】また、別の問題もある。図41は、従来の表の挿入処理を説明するための第3の例の説明図、図42は、従来の表の挿入処理の第3の例において章の追加を行なう場合の説明図である。図41に示すように、入力文書中に「表3」が追加されたが、テンプレート文書は変更されず、不整合が生じた場合を考える。この場合、入力文書の構造は図39と共通であり、テンプレート文書の構造は図37と共通である。入力文書中の「表1」、「表2」、「表3」には、属性値「挿入部品」が付与されており、テンプレート文書中の2つの「本文」には、属性値「末子に挿入」が付与されているとする。この場合、「表3」は、挿入すべき章がテンプレート中

に存在せず、挿入処理はエラーとなる。ユーザには、図42に示すように、入力文書中の表の数に併せて「第3章」を新設するような処理が望まれるが、従来の技術では、これを解決する有効な手段を示しておらず、エラーとして例外処理を行なうことになる。そのため、入力文書中の表の数に合わせて第3章を新設するような処理は実現できない。

【0011】このように、従来の技術では、文書部品群に対して一括した文書処理を行なうことは可能であるが、処理対象の文書が変更されるとユーザに煩雑な作業を強いることになる。また、処理内容が処理対象の文書の構造その他によって切り替わるような処理を構築することができない。

【0012】

【発明が解決しようとする課題】本発明は、上述した事情に鑑みてなされたもので、ユーザに繁雑な作業を強いることなく、構造化文書の複数の文書部品に対して一括した文書処理を行なうことのできる構造化文書処理装置を提供することを目的とするものである。

【0013】

【課題を解決するための手段】本発明は、構造化文書に対して処理を行なう構造化文書処理装置において、入力文書から所定の条件に従って部分構造を抜き出しこれに基づき部分構造列を生成する部分構造列抽出手段と、該部分構造列抽出手段により生成された前記部分構造列を用いて指示された処理を行ない出力文書を作成する処理手段を有することを特徴とするものである。前記部分構造列抽出手段は、複数の入力文書から複数の部分構造列を抽出し、保持するように構成することができる。

【0014】また、この構造化文書処理装置は、出力文書のひな型となるテンプレート文書を保持するテンプレート文書保持手段を有し、前記処理手段は、前記部分構造列抽出手段により生成された前記部分構造列を用い前記テンプレート文書保持手段に保持されているテンプレート文書に基づき指示された処理を行ない出力文書を作成することを特徴とするものである。

【0015】さらに、前記テンプレート文書保持手段を有する構造化文書処理装置において、処理の指示あるいは前記テンプレート文書保持手段に保持されている前記テンプレート文書の情報に従い、前記部分構造列抽出手段により生成された前記部分構造列に対して処理を行なう部分構造列処理手段を有することを特徴とするものである。

【0016】また、前記テンプレート文書保持手段を有する構造化文書処理装置において、前記テンプレート文書保持手段に保持されている前記テンプレート文書に対して、処理の指示あるいは前記部分構造列抽出手段により生成された前記部分構造列に基づいた処理を行なうテンプレート文書処理手段を有することを特徴とするものである。

## 【0017】

【作用】本発明によれば、入力された構造化文書は、部分構造列抽出手段によって、処理に必要な文書部品が抽出され、部分構造列が生成される。処理手段では、生成された部分構造列を一括して処理対象として扱い、指示された構造化文書処理を行なう。その結果を出力文書として出力する。また、テンプレート文書保持手段を設けてテンプレート文書を保持しておき、処理手段において、部分構造列を用いてテンプレート文書に対して処理を行ない、出力文書を作成する。

【0018】これにより、処理に必要な各文書部品は、ある条件をもとに抽出し、ある場合には、抽出された文書部品の列に対しての文書部品の抽出や処理等を行なって、部分構造列を生成しているので、従来のように、それぞれに属性値を付与することなく、必要な文書部品を抽出することができ、ユーザの利便性を向上させることができる。

【0019】また、部分構造列処理手段において、指示された条件あるいはテンプレート文書の情報に従って、部分構造列に対して、さらに部分列の抽出など、種々の処理を行なうことができる。例えば、部分構造列とテンプレート文書が不整合であっても、テンプレート文書に合わせた処理を行なって、出力文書を得ることができる。さらに、テンプレート文書処理手段において、部分構造列に基づいてテンプレート文書に対して処理を行なうことができるので、例えば、部分構造列とテンプレート文書が不整合の場合に、部分構造列にテンプレート文書を整合させて処理を行ない、出力文書を得ることができる。このため、従来のように不整合のためにエラーとなることなく、所望の出力文書を得ることができる。

## 【0020】

【実施例】図1は、本発明の構造化文書処理装置の第1の実施例を示すブロック構成図である。図中、1は入力文書、2は入力文書保持部、3は部分構造列抽出部、4は抽出指示部、5は処理実行部、6は処理指示部、7は出力文書である。

【0021】入力文書保持部2は、構造化文書である入力文書1を保持し、部分構造列抽出部3に対して、入力文書1の構造および内容を引き渡す。部分構造列抽出部3は、抽出指示部4からの抽出指示に基づいて、入力文書1の構造から、部分構造を抜き出し、部分構造列を生成する。処理実行部5は、部分構造列抽出部3において生成された部分構造列、及び、処理指示部6の処理指示に基づいて、構造化文書処理を実行し、その結果を出力文書7として出力する。出力する出力文書7は、構造化文書の形式で出力しても良いし、文書のイメージとして出力することも可能である。抽出指示部4及び処理指示部6は、それぞれ、抽出指示、処理指示を保持しておく構成としても良い。

## 【0022】本発明の構造化文書処理装置の第1の実施

例における動作を具体例をもとに説明する。図2は、本発明の構造化文書処理装置の第1の実施例における具体的な処理例の説明図、図3は、図2における入力文書の論理構造の説明図、図4は、図2における入力文書から

- 05 抽出された部分構造列の説明図である。図2に示すように、入力文書は、2つの章を有している。入力文書中の章題を抽出して目次のような一覧を作成する処理を行なう。入力文書の論理構造を図3に示す。入力文書のイメージを図2に示してはいるが、実際には、入力文書1は
- 10 図3に示す論理構造の形式で入力文書保持部2に保持される。入力文書から抽出される部分構造列を図4に示す。部分構造列は、図3に示す入力文書の論理構造から、処理に必要な要素である章が2つ順序を持って並べられている。図中の矢印は、その順序の方向を示している。
- 15 部分構造列としては、章題のほか、本文や、段落、表、図、これらを含めた章や節全体、あるいは文書全体等、種々の構造をその要素とすることができる。

- 【0023】図3に示す入力文書1が入力文書保持部2に保持されている。抽出指示部4からは、「入力文書から章、章題を抽出し、部分構造列とする」という指示が部分構造列抽出部3に送られる。部分構造列抽出部3において、入力文書1から図4に示した部分構造列を抜き出す。

- 【0024】一方、処理指示部6からは、「部分構造列の各要素を列記する」という指示が、処理実行部5に送られる。処理実行部5において、部分構造列抽出部3において生成された部分構造列をもとに、章題を列記した文書を生成し、図2に示すような出力文書として出力する。

- 【0025】このように、この実施例では、抽出指示に従って部分構造列の抽出、生成を行なっているので、例えば、入力文書1中に章の数が増加した場合であっても、同じ動作を行なわせることによって、所望の文書を出力文書として得ることができる。このとき、従来のように、増加した章に属性値をいちいち付与する必要はない。

- 【0026】図5は、本発明の構造化文書処理装置の第2の実施例を示すブロック構成図である。図中、図1と同様の部分には同じ符号を付して説明を省略する。8は部分構造列管理部、9はテンプレート文書保持部、10はテンプレート文書である。この実施例では、部分構造列管理部8及びテンプレート文書保持部9を有する例を示している。部分構造列抽出部3において抜き出され、生成された部分構造列は、部分構造列管理部8において保持、管理される。部分構造列管理部8は、その内部に、例えば、部分構造列のデータベースを有する構成とすることができる。出力文書7のひな型であるテンプレート文書10は、テンプレート文書保持部9に保持される。処理実行部5は、処理指示部6から送られる処理指示に従って、テンプレート文書保持部9からテンプレー

ト文書10を、また、部分構造列管理部8から部分構造列を得て文書処理を実行する。その結果を出力文書7として出力する。

【0027】この実施例では、部分構造列管理部8を有しているので、複数の部分構造列および複数文書の部分構造列を保持、管理することができる。部分構造列管理手段8は、上述のように、その内部に部分構造列のデータベースを有する構成の他、部分構造列のデータベースを持たずに、部分構造列のデータベースに対するインターフェースとしても実現可能である。データベースでなくとも、単に部分構造列を保持できる構成であれば、実現することができる。もちろん、図1に示したように、部分構造列管理手段8を有しない構成も可能である。

【0028】本発明の構造化文書処理装置の第2の実施例における動作を具体例をもとに説明する。図6は、本発明の構造化文書処理装置の第2の実施例における第1の具体的な処理例の説明図、図7は、図6における入力文書の論理構造の説明図、図8は、図6におけるテンプレート文書とその論理構造の説明図、図9は、図6における入力文書から抽出された部分構造列の説明図、図10(A)は、図6における出力文書、図10(B)は、その論理構造の説明図である。

【0029】図6に示すように、入力文書は、2つの表と2つの段落とからなり、テンプレート文書は、2つの章と各章に属する段落とからなる。入力文書中の表をテンプレート文書中の各章の末尾に挿入する処理を行う。入力文書の構造を図7に、テンプレート文書の構造を図8に示す。入力文書から抽出される部分構造列を図9に示す。この例の場合、抽出される文書部品は、2つの表である。

【0030】図7に示す入力文書は入力文書保持部2に、また、図8に示すテンプレート文書はテンプレート文書保持部9に、それぞれ保持される。抽出指示部4からは、例えば、「入力文書から図表を抽出し、部分構造列とする」という指示が部分構造列抽出部3に送られる。部分構造列抽出部3では、入力文書1から抽出指示に従って、図7において点線で囲んで示す部分構造を抜き出し、図9に示す部分構造列を生成する。生成された部分構造列は、部分構造列管理部8に送られ、保持、管理される。

【0031】一方、処理指示部6からは、例えば、「部分構造列の各要素をテンプレート文書の各章の章末に挿入」という指示が、処理実行部5に送られる。処理実行部5において、処理指示に従い、部分構造列管理部8に対して部分構造列を要求し、部分構造列管理部8から要求した部分構造列を受け取るとともに、テンプレート文書保持部9に保持されているテンプレート文書に対し、部分構造列の挿入処理を行なう。そして、処理結果を出力文書7として出力する。この処理により、出力文書は、図10に示すように、各章のおわりに入力文書中の

表が挿入されている。

【0032】挿入すべき表の数が増加した場合であっても、挿入される側の文書の章の数が対応していれば、同じ動作によって所望の文書を出力文書として得ることができる。図11は、本発明の構造化文書処理装置の第2の実施例における第2の具体的な処理例の説明図、図12は、図11における入力文書の論理構造の説明図である。図13(A)は、図11におけるテンプレート文書、図13(B)は、その論理構造の説明図である。図14は、図11における入力文書から抽出された部分構造列の説明図、図15は、図11における出力文書の論理構造の説明図である。

【0033】図11に示すように、入力文書中の表の数が3つに増加し、同時にテンプレート文書の章の数も3つに増加したとする。このとき、挿入すべき表の数も3つに増加した。しかし、抽出指示部4から送られる「入力文書から図表を抽出し、部分構造列とする」という指示、および、処理指示部6から送られる「部分構造列の各要素をテンプレート文書の各章の章末に挿入」という指示が実行されることによって、第1の具体的な処理例と同様の手順で所望の文書が得られる。これらの指示は、第1の具体的な処理例と同じである。すなわち、図12に示した入力文書から、点線で囲んだ3つの表を抜き出し、図14に示すような部分構造列を生成する。そして、図13に示すテンプレート文書に対して、生成した部分構造列の各表を各章の章末に挿入処理し、図15に示す論理構造の出力文書を生成する。

【0034】また、一部の表が図に置き換わった場合であっても同じ動作によって所望の文書を出力文書として得ることができる。図16は、本発明の構造化文書処理装置の第2の実施例における第3の具体的な処理例の説明図、図17は、図16における入力文書の論理構造の説明図、図18は、図16における入力文書から抽出された部分構造列の説明図、図19(A)は、図16における出力文書、図19(B)は、その論理構造の説明図である。

【0035】図16に示した入力文書、テンプレート文書は、図6に示した入力文書、テンプレート文書に比べて「表2」が「図1」と置き換わっている。しかし、抽出指示部4から送られる「入力文書から図表を抽出し、部分構造列とする」という指示、および、処理指示部6から送られる「部分構造列の各要素をテンプレート文書の各章の章末に挿入」という指示が実行されることにより、上述の第1の具体的な処理例と同様の指示、手順で、所望の文書が得られる。すなわち、図17に示す入力文書から、点線で囲んだ表と図が抜き出され、図18に示す部分構造列が生成される。テンプレート文書は、図8に示されたものと同じである。生成された部分構造列は、テンプレート文書の各章の章末にそれぞれ挿入され、図19に示すような出力文書を得ることができる。

【0036】このように、第2の実施例では、テンプレート文書への挿入処理において、入力文書の文書部品が増加したり、変更されたりしても、新たに属性値を付与することなく、同じ指示により所望の出力文書を得ることができる。

【0037】また、図20に示すような複数の文書からの図表の挿入の処理を行なうこともできる。図20は、本発明の構造化文書処理装置の第2の実施例における第4の具体的な処理例の説明図、図21は、図20における入力文書Bの論理構造の説明図、図22は、図20における入力文書から抽出された部分構造列の説明図、図23(A)は、図20における出力文書、図23(B)は、その論理構造の説明図である。は、図20における出力文書とその論理構造の説明図である。

【0038】入力文書Aは、2つの段落と2つの表からなり、入力文書Bは2つの段落と2つの図からなる。テンプレート文書は、2つの章からなる。入力文書Aから2つの表を抜き出し、テンプレート文書の各章の末尾に挿入し、かつ、入力文書Bから2つの図を抜き出し、テンプレート文書の各章の先頭に挿入する処理を行なう。入力文書Bの構造を図21に示す。入力文書Aの構造は、図7に示されたものと同じである。テンプレート文書は、図8に示されたものと同じである。図22に2つの入力文書から抽出され、生成された2つの部分構造列を示す。

【0039】図7に示す入力文書Aおよび図21に示す入力文書Bを入力文書保持部2に、また、図8に示すテンプレート文書をテンプレート文書保持部9にそれぞれ保持する。抽出指示部4からは、「入力文書から図表を抽出し、部分構造列とする」という指示が、部分構造列抽出部3に送られる。

【0040】部分構造列抽出部3において、図7に示す入力文書Aから、点線で囲んだ2つの表を抜き出し、部分構造列Aを作成する。また、図21に示す入力文書Bから、点線で囲んだ2つの図を抜き出し、部分構造列Bを作成する。部分構造列A、Bは、部分構造列抽出部3から部分構造列管理部8に送られる。部分構造列管理部8は、部分構造列Aを例えば「streamA」と、部分構造列Bを例えば「streamB」と名付けて、それぞれ保持、管理する。

【0041】一方、処理保持部6から、例えば、「『streamA』の各要素をテンプレート文書の各章の章末に挿入」および「『streamB』の各要素をテンプレート文書の各章の先頭に挿入」という指示が処理実行部5に送られる。処理実行部5は、「streamA」を部分構造列管理部8に要求し、送られてきた「streamA」をテンプレート文書10の各章の章末に挿入する。さらに、処理実行部5は、「streamB」を部分構造列管理部8に要求し、送られてきた「streamB」を、「streamA」が挿入された後

のテンプレート文書の各章の先頭に挿入する。そして、「streamA」、「streamB」が挿入された後の処理結果を、図23に示すような出力文書として出力する。

- 05 【0042】図と表の位置が入れ替わった出力文書であっても、簡単に作成することができる。図24(A)は、図20における図と表が入れ替わった出力文書、図24(B)は、その論理構造の説明図である。作成方法としては、入力文書Aと入力文書Bを入れ替えるだけでよく、処理方法をなんら変更せずに得ることができる。もちろん、入力文書Aと入力文書Bをそのままにし、処理指示を「『streamA』の各要素をテンプレート文書の各章の先頭に挿入」および「『streamB』の各要素をテンプレート文書の各章の章末に挿入」という指示とすることもできる。

【0043】図25は、本発明の構造化文書処理装置の第3の実施例を示すブロック構成図である。図中、図5と同様の部分には同じ符号を付して説明を省略する。11は部分構造列処理部、12は部分構造列処理指示部である。

【0044】部分構造列処理部11は、部分構造列処理指示部12から送られる部分構造列処理指示に基づいて部分構造列管理手段に保持されている部分構造列を処理する。例えば、部分構造列から、部分構造列処理指示で指示された条件に応じて必要な部分構造列を抜き出すことができる。また、テンプレート文書保持部9に保持されているテンプレート文書10の構造や内容を参照し、これらの情報に基づいて部分構造列に対して処理を行なうことができる。処理された部分構造列は、再び部分構造列管理手段8に保持、管理される。部分構造列処理指示部12は、部分構造列処理指示を保持しておく機能を有するように構成することもできる。

【0045】本発明の構造化文書処理装置の第3の実施例における動作を具体例をもとに説明する。図26は、本発明の構造化文書処理装置の第3の実施例における第1の具体的な処理例の説明図、図27は、本発明の構造化文書処理装置の第3の実施例における第2の具体的な処理例の説明図、図28は、図26、図27における入力文書の論理構造の説明図、図29は、図26、図27における入力文書から抽出された部分構造列の説明図、

図30は、図27におけるテンプレート文書Bを用いたときの出力文書の論理構造の説明図である。

【0046】図26、図27に示すように、複数の文書から図表を抽出し、挿入の処理を行なうこととする。図26においては、入力文書は、3つの段落、2つの表、1つの図からなり、テンプレート文書Aは、2つの章からなる。このとき、入力文書中の2つの表を、テンプレート文書Aの各章末に挿入する。図27においては、入力文書は図26と同一のものであるが、テンプレート文書Bは、3つの章から構成されている。このとき、入力

文書中の2つの表と1つの図をテンプレート文書Bの各章末に挿入する。すなわち、図26に示すように、テンプレート文書が2章構成であれば、表のみを入力文書から抜き出してテンプレート文書の各章末に挿入し、それ以外の場合には、入力文書中の全ての図表を抜き出してテンプレート文書の各章末に挿入するものとする。

【0047】入力文書の構造を図28に示している。テンプレート文書Aは、図8に示されたものと同じものである。テンプレート文書Bは、図13に示されたものと同じものである。図29において、部分構造列Aは入力文書から抽出した部分構造列を示し、部分構造列Bは、部分構造列処理部11による処理後の部分構造列を示す。テンプレート文書Aによる出力文書は、図10に示したものと同じものである。図30にテンプレート文書Bによる出力文書の構造を示す。

【0048】まず、図28に示す入力文書を入力文書保持部2に、図8または図13に示すテンプレート文書をテンプレート文書保持部9にそれぞれ保持する。抽出指示部4から、例えば、「入力文書から図表を抽出し、部分構造列とする」という指示が、部分構造列抽出部3に送られる。部分構造列抽出部3において、入力文書から図29に点線で囲んで示す2つの表と1つの図を抜き出し、部分構造列Aを作成する。部分構造列Aは、部分構造列抽出部3から部分構造列管理部8に送られる。部分構造列管理部8は、部分構造列Aを保持、管理する。

【0049】部分構造列処理指示部12から部分構造列処理部11へ、例えば、「テンプレート文書中の章の数が2であれば部分構造列から表だけを抜き出したものを新たに部分構造列とする。」という部分構造列処理指示が送られる。部分構造列処理部11は、テンプレート文書保持部9に保持されているテンプレート文書10の構造を調べる。その結果、章の数が2でなければ何も実行しない。また、章の数が2であれば、部分構造列管理部8に部分構造列を要求し、送られてきた部分構造列Aから表のみを抜き出し、新しく図29に示す部分構造列Bを作成する。部分構造列Bは、部分構造列管理部8に送られる。部分構造管理手段8は、部分構造列Aを削除し、部分構造列Bを部分構造列として保持、管理する。

【0050】一方、処理指示部6から、例えば、「部分構造列の各要素をテンプレート文書の各章の章末に挿入」という指示が処理実行部5に送られる。処理実行部5は、部分構造列を部分構造列管理部8に要求し、送られてきた部分構造列をテンプレート文書の各章の章末に挿入する。挿入された後の処理結果を、図10あるいは図30に示すような出力文書として出力する。

【0051】このように、部分構造列処理部11により、用いるテンプレート文書に従って部分構造列を処理することができるので、処理指示を変更せずに、テンプレート文書に従った出力文書を得ることができる。

【0052】上述の処理の中で、部分構造列処理部11

において作成された部分構造列Bにより、もとの部分構造列Aを削除しているが、これを削除せず、テンプレート文書の種類とともに、部分構造列A及び部分構造列Bをともに保持、管理することもできる。このとき、処理指示部6で使用することが指定されたテンプレート文書に従って、部分構造列管理部8が部分構造列を選択し、処理実行部へ送るように構成することができる。

【0053】図31は、本発明の構造化文書処理装置の第4の実施例を示すブロック構成図である。図中、図210と同様の部分には同じ符号を付して説明を省略する。13はテンプレート文書処理部、14はテンプレート文書処理指示部である。

【0054】テンプレート文書処理部13は、テンプレート文書の前処理として、テンプレート文書処理指示部1514の指示に従い、テンプレート文書の処理を行なう。その際に、部分構造列管理部8に保持されている部分構造列の要素の数、内容等、部分構造列に関する情報を参照にすることもできる。処理指示部6に保持された指示に従って、処理実行部5は、テンプレート文書処理部1203で前処理されたテンプレート文書と、部分構造列管理部8から部分構造列を得て文書処理を実行することになる。

【0055】図32は、本発明の構造化文書処理装置の第4の実施例における具体的な処理例の説明図、図3325は、図32における入力文書から抽出された部分構造列の説明図、図34は、図32における出力文書の論理構造の説明図である。この例では、図32に示すような図表の挿入の処理を行なうこととする。すなわち、入力文書中から図表を抜き出し、テンプレート文書中の各章の章末に挿入する。ただし、入力文書中の図表の数がテンプレート文書中の章の数より大きい場合、超過している数だけの章をテンプレート文書中に追加した上で図表の挿入を行なう。図32においては、入力文書は、3つの段落、2つの表、1つの図からなり、テンプレート文書は2つの章からなる。入力文書中の表1、図35を第1章、第2章の各章末に挿入する。表2は、テンプレート文書に章題のみを持つ第3章を追加した上で第3章に挿入する。

【0056】入力文書の構造は、図28に示されたものと同じである。テンプレート文書は、図8に示されたものと同じものである。図33に抽出される部分構造列を示す。出力文書を図34に示す。図32及び図34中、一点鎖線で囲まれた部分は、自動的に追加された第3章の部分である。

【0057】まず、図28に示す入力文書を入力文書保持部2に、また、図8に示すテンプレート文書をテンプレート文書保持部9にそれぞれ保持する。抽出指示部4から、例えば、「入力文書から図表を抽出し、部分構造列とする」という指示が、部分構造列抽出部3に送られる。部分構造列抽出部3において、図28に示した入力

文書から、点線で囲んだ図、表を抜き出し、図33に示す部分構造列を作成する。作成された部分構造列は、部分構造列抽出部3から部分構造列管理部8に送られる。部分構造列管理部8は、部分構造列を保持、管理する。この例では、部分構造処理指示部、部分構造列処理部は動作しない。

【0058】テンプレート文書処理指示部14から、例えば、「部分構造列の要素の数とテンプレート文書の章の数と比べて、部分構造列の要素の数が大きい場合は、要素の数と章の数が同じになるようにテンプレート文書に章を追加する。それ以外の場合は、テンプレート文書を加工しない。」という指示をテンプレート文書処理部13に送る。テンプレート文書処理部13は、送られた指示を実行する。図28に示した入力文書の例の場合、図33に示すように、部分構造列の要素の数は3であり、図8に示すテンプレート文書の章の数は2である。よって、テンプレート文書処理部13は、章を1つテンプレート文書に追加する。

【0059】処理保持部6から、例えば、「部分構造列の各要素をテンプレート文書の各章の章末に挿入」という処理指示が処理実行部5に送られる。処理実行部5は、部分構造列を部分構造列管理部8に要求し、送られてきた部分構造列をテンプレート文書の各章の章末に挿入する。このとき用いられるテンプレート文書は、テンプレート文書処理部13において章の追加されたものが用いられる。その結果、図34に示すように、第3章が追加され、各章に図表が挿入された出力文書が出力される。

【0060】このように、入力文書とテンプレート文書の不整合の場合であっても、処理指示を変更することなく、自動的に整合を図って文書処理を行なうことができる。この例において、テンプレート文書の形式を優先させ、章を追加せずに挿入処理を行なう場合には、部分構造処理指示部12から、テンプレート文書に合わせた部分構造列を作成するように指示することにより、所望の処理を行なわせることができる。

【0061】上述の各実施例において、各部はそれぞれ別のモジュールで構成され、実行されるので、全体に影響を与えることなく、部分の定義を変更することが可能となっている。このことによって高い保守性と拡張性が実現されている。

【0062】上述の各実施例において、処理実行部5、部分構造列処理部11、テンプレート文書処理部13には、例えば、特願平4-206722号に記載されている処理方法または処理装置を用いて構築することができる。また、本発明の構造化文書処理装置は、具体的には、文書処理機能を有する装置、例えばワークステーションやコンピュータなどによって実現することができる。

【0063】

【発明の効果】以上の説明から明らかなように、本発明によれば、ユーザに対して過度の負担をかけることなく、構造化文書の複数の文書部品を一括して扱う処理が可能となり、ユーザの利便性を向上させることができ

- 05 る。
- 【0064】また、例えば、入力文書から抽出した部分構造列とテンプレート文書が不整合であっても、部分構造列あるいはテンプレート文書に合わせた処理を行なって、出力文書を得ることができる。このため、従来のように不整合のためにエラーとなることなく、所望の出力文書を得ることができるという効果がある。
- 【図面の簡単な説明】
- 【図1】 本発明の構造化文書処理装置の第1の実施例を示すブロック構成図である。
- 10 【図2】 本発明の構造化文書処理装置の第1の実施例における具体的な処理例の説明図である。
- 【図3】 図2における入力文書の論理構造の説明図である。
- 15 【図4】 図2における入力文書から抽出された部分構造列の説明図である。
- 【図5】 本発明の構造化文書処理装置の第2の実施例を示すブロック構成図である。
- 【図6】 本発明の構造化文書処理装置の第2の実施例における第1の具体的な処理例の説明図である。
- 20 【図7】 図6における入力文書の論理構造の説明図である。
- 【図8】 図6におけるテンプレート文書とその論理構造の説明図である。
- 【図9】 図6における入力文書から抽出された部分構造列の説明図である。
- 25 【図10】 図6における出力文書とその論理構造の説明図である。
- 【図11】 本発明の構造化文書処理装置の第2の実施例における第2の具体的な処理例の説明図である。
- 【図12】 図11における入力文書の論理構造の説明図である。
- 30 【図13】 図11におけるテンプレート文書とその論理構造の説明図である。
- 【図14】 図11における入力文書から抽出された部分構造列の説明図である。
- 35 【図15】 図11における出力文書とその論理構造の説明図である。
- 【図16】 本発明の構造化文書処理装置の第2の実施例における第3の具体的な処理例の説明図である。
- 40 【図17】 図16における入力文書の論理構造の説明図である。
- 【図18】 図16における入力文書から抽出された部分構造列の説明図である。
- 45 【図19】 図16における出力文書とその論理構造の説明図である。
- 50 【図20】 図16における出力文書とその論理構造の説明図である。

# 構造化文書処理装置

特開平7-56920

【図20】 本発明の構造化文書処理装置の第2の実施例における第4の具体的な処理例の説明図である。

【図21】 図20における入力文書Bの論理構造の説明図である。

【図22】 図20における入力文書から抽出された部分構造列の説明図である。

【図23】 図20における出力文書とその論理構造の説明図である。

【図24】 図20における図と表が入れ替わった出力文書とその論理構造の説明図である。

【図25】 本発明の構造化文書処理装置の第3の実施例を示すブロック構成図である。

【図26】 本発明の構造化文書処理装置の第3の実施例における第1の具体的な処理例の説明図である。

【図27】 本発明の構造化文書処理装置の第3の実施例における第2の具体的な処理例の説明図である。

【図28】 図26、図27における入力文書の論理構造の説明図である。

【図29】 図26、図27における入力文書から抽出された部分構造列の説明図である。

【図30】 図27におけるテンプレート文書Bを用いたときの出力文書の論理構造の説明図である。

【図31】 本発明の構造化文書処理装置の第4の実施例を示すブロック構成図である。

【図32】 本発明の構造化文書処理装置の第4の実施例における具体的な処理例の説明図である。

【図33】 図32における入力文書から抽出された部

分構造列の説明図である。

【図34】 図32における出力文書の論理構造の説明図である。

【図35】 従来の表の挿入処理を説明するための第1例の説明図である。

【図36】 図35における入力文書の論理構造の説明図である。

【図37】 図35におけるテンプレート文書の論理構造の説明図である。

10 【図38】 従来の表の挿入処理を説明するための第2例の説明図である。

【図39】 図38における入力文書の論理構造の説明図である。

15 【図40】 図38におけるテンプレート文書の論理構造の説明図である。

【図41】 従来の表の挿入処理を説明するための第3例の説明図である。

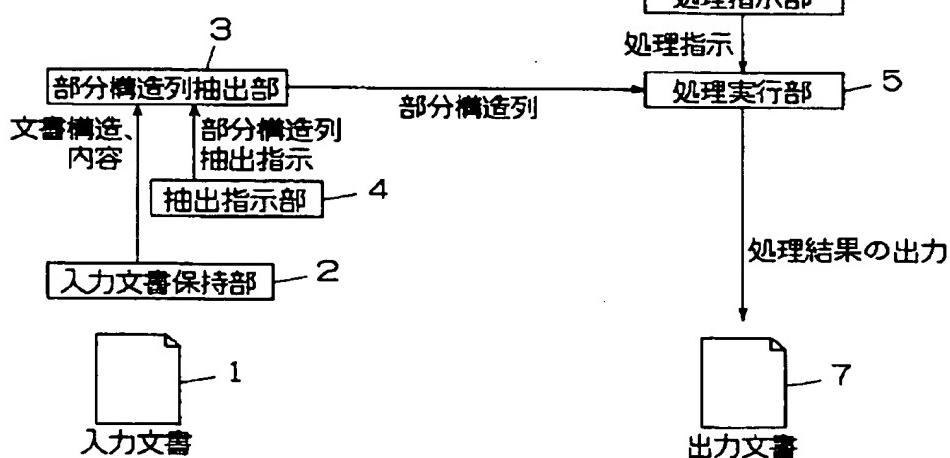
【図42】 従来の表の挿入処理の第3の例において章の追加を行なう場合の説明図である。

20 【符号の説明】

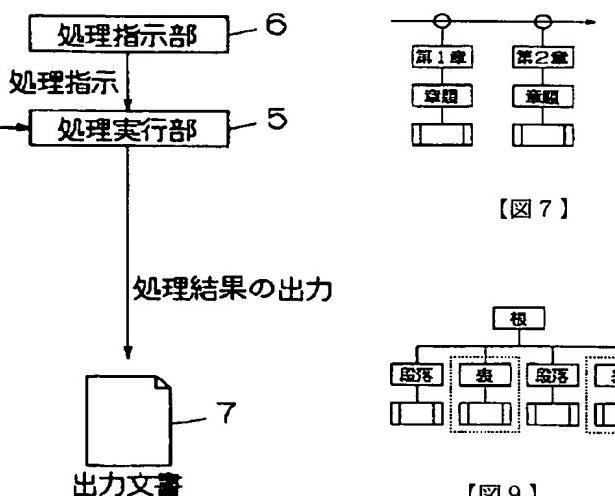
1 入力文書、2 入力文書保持部、3 部分構造列抽出部、4 抽出指示部、5 処理実行部、6 処理指示部、7 出力文書、8 部分構造列管理部、9 テンプレート文書保持部、10 テンプレート文書、11 部分

25 構造列処理部、12 部分構造列処理指示部、13 テンプレート文書処理部、14 テンプレート文書処理指示部。

【図1】



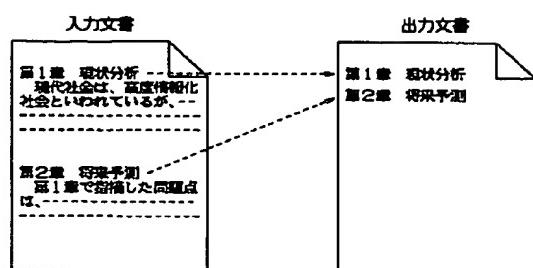
【図4】



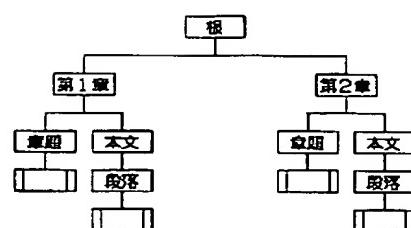
# 構造化文書処理装置

特開平7-56920

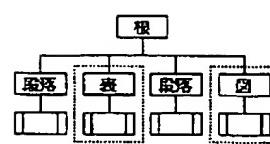
【図2】



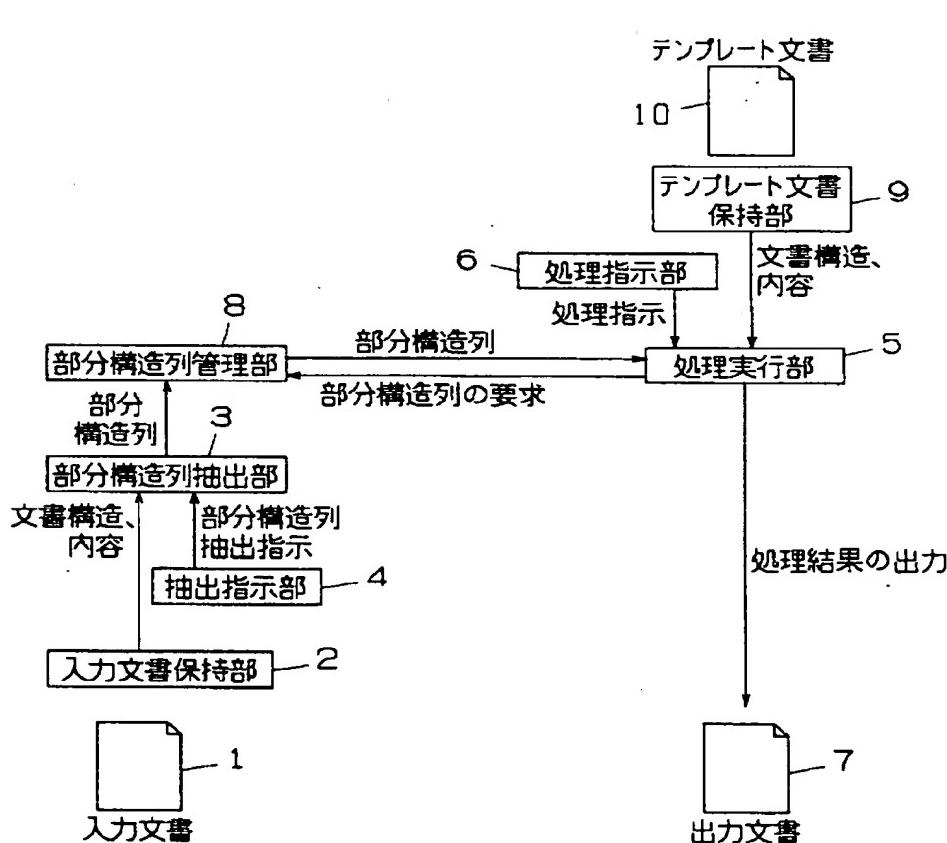
【図3】



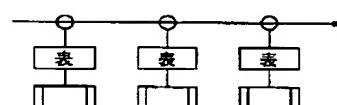
【図17】



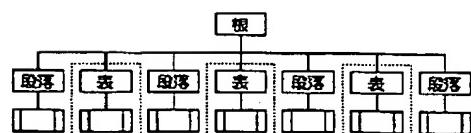
【図5】



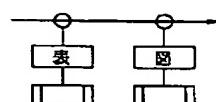
【図14】



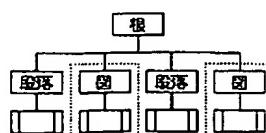
【図12】



【図18】



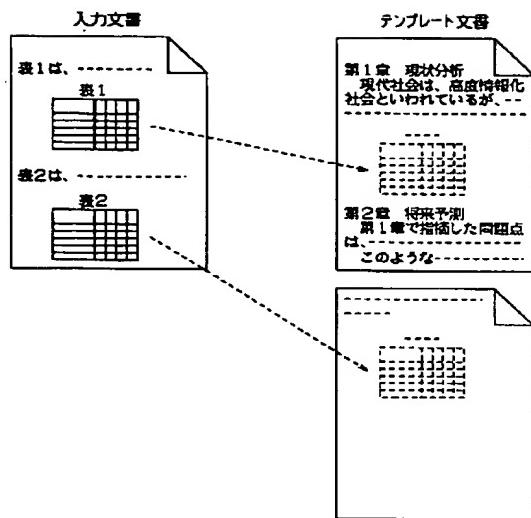
【図21】



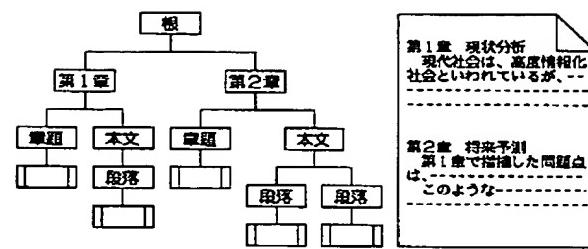
# 構造化文書処理装置

特開平7-56920

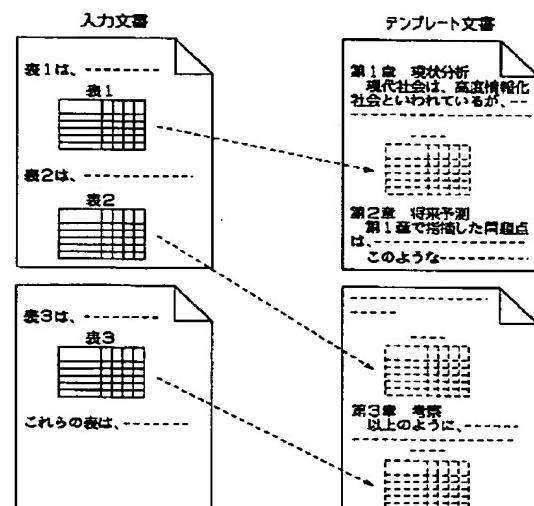
【図 6】



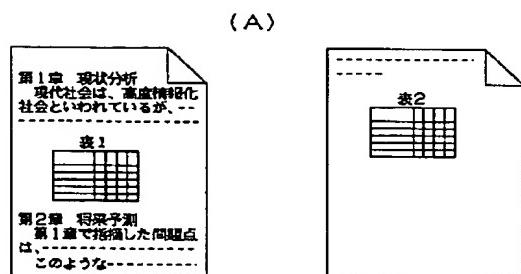
【図 10】



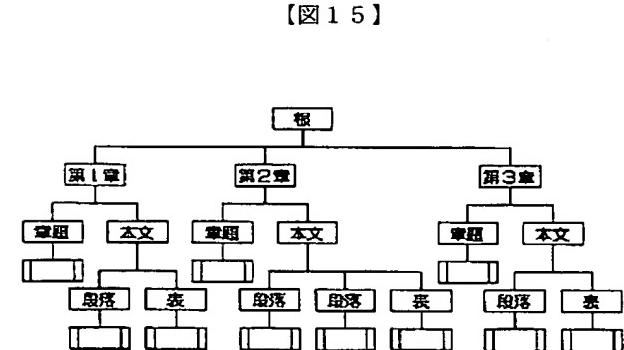
【図 8】



【図 11】

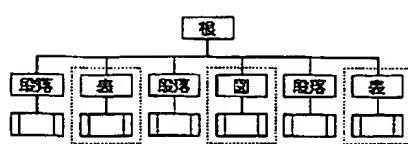
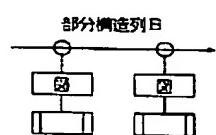
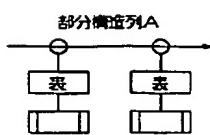


(A)



【図 15】

【図 22】

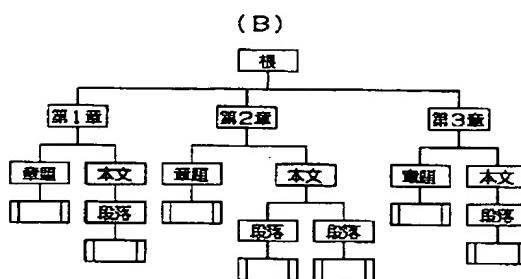
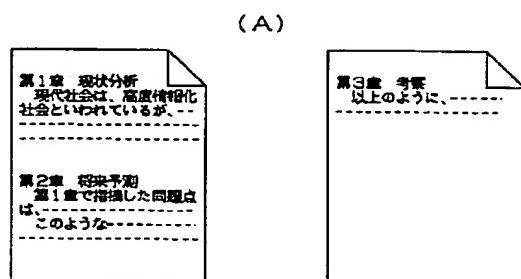


【図 28】

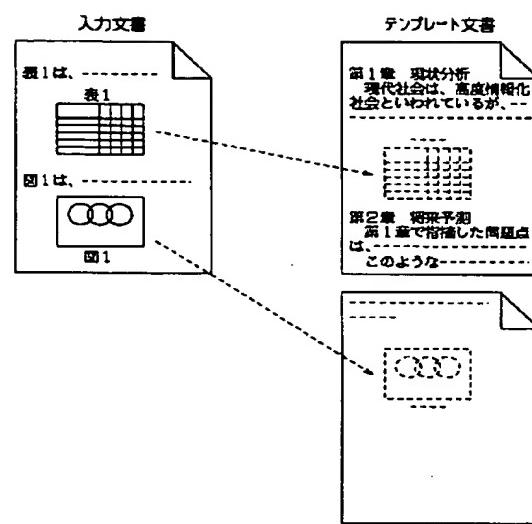
# 構造化文書処理装置

特開平7-56920

【図13】

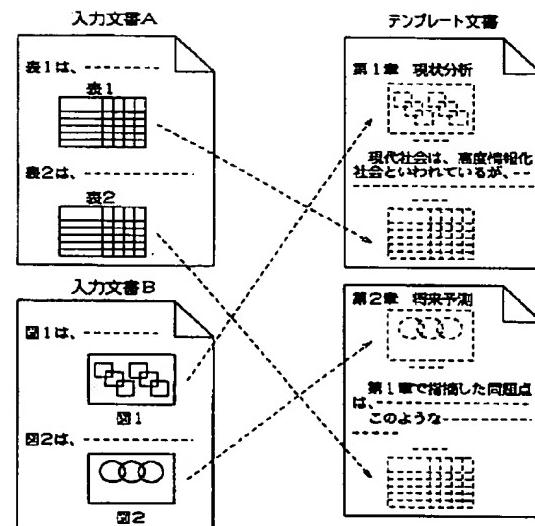
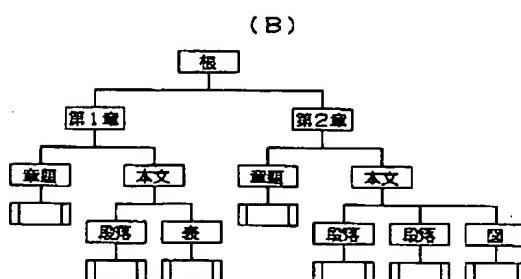
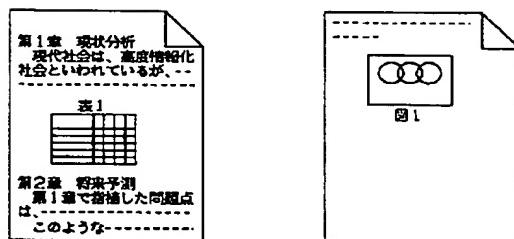


【図16】

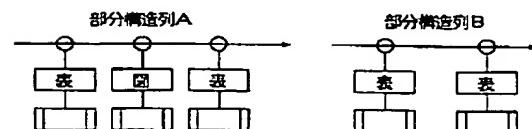


【図20】

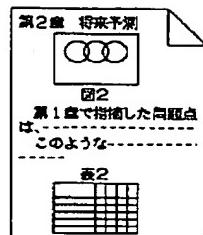
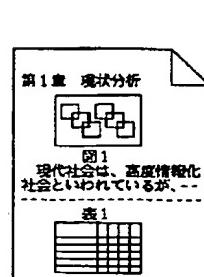
【図19】



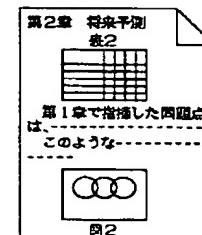
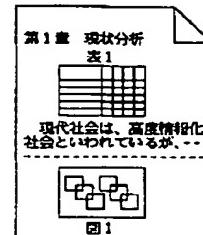
【図29】



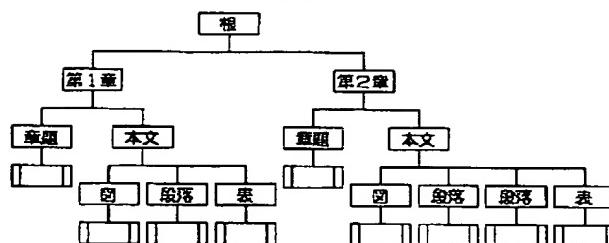
【図23】



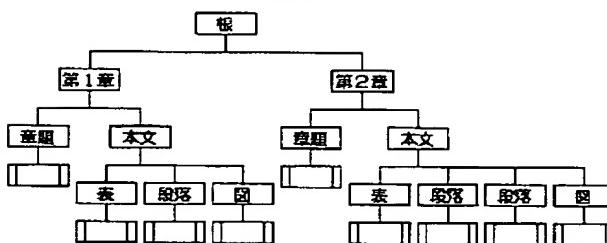
【図24】



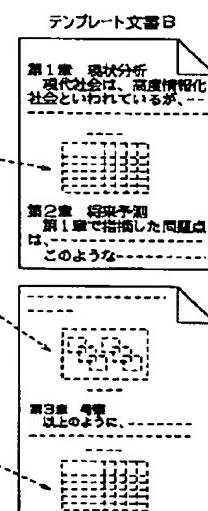
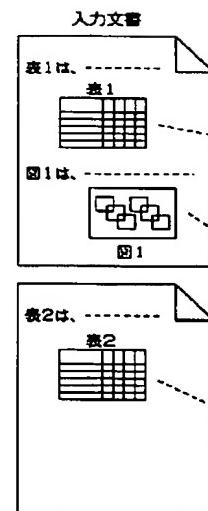
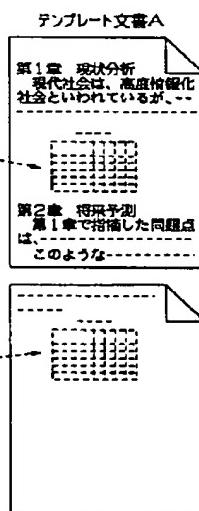
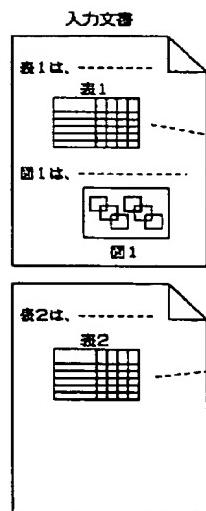
(B)



(B)



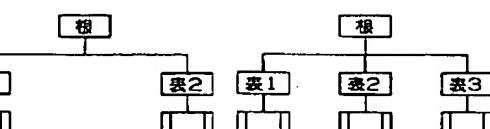
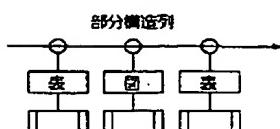
【図26】



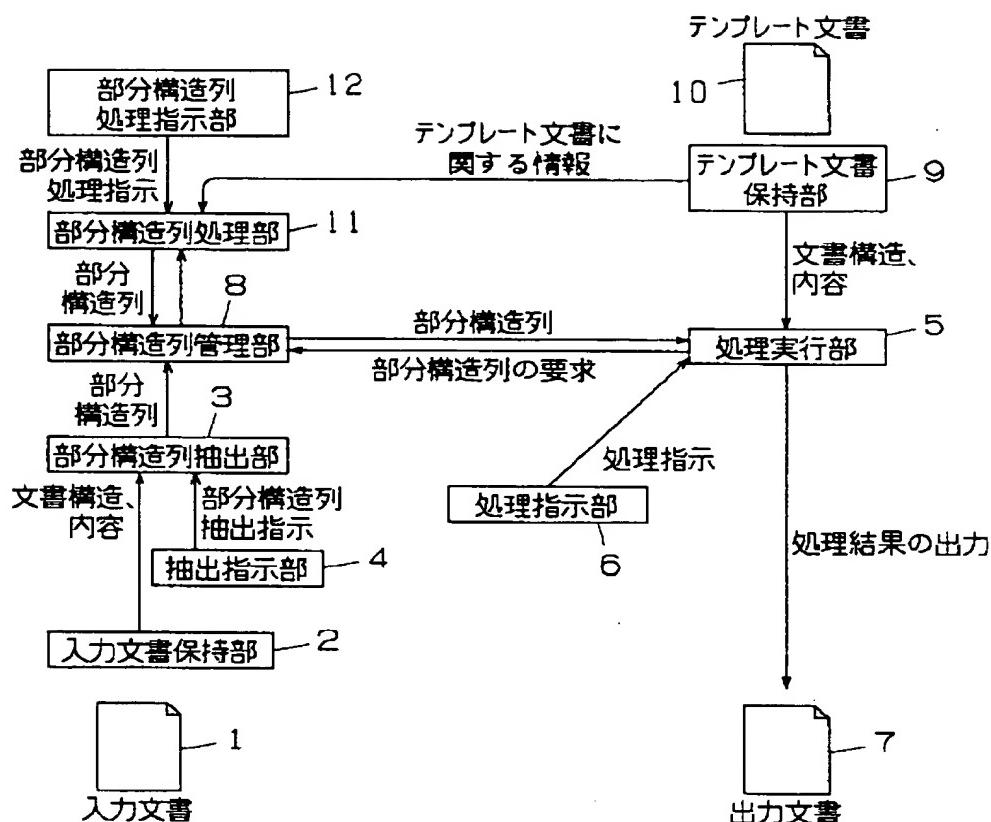
【図33】

【図36】

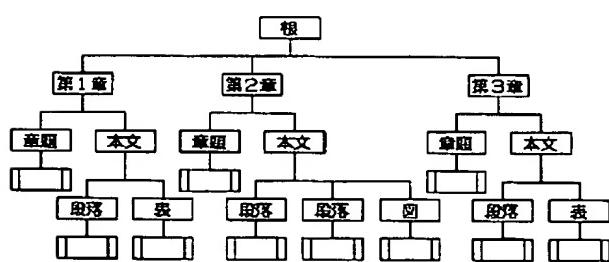
【図39】



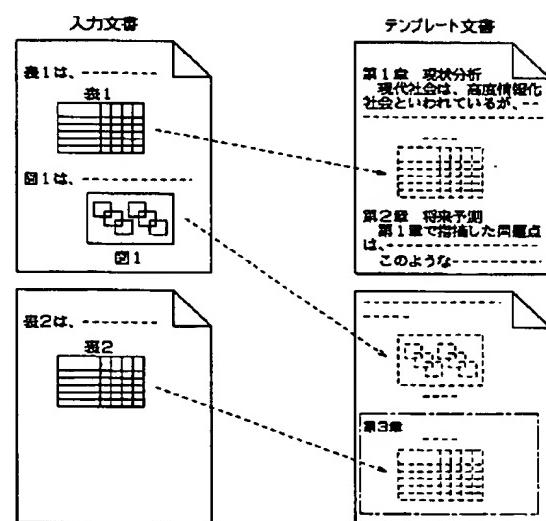
【図25】



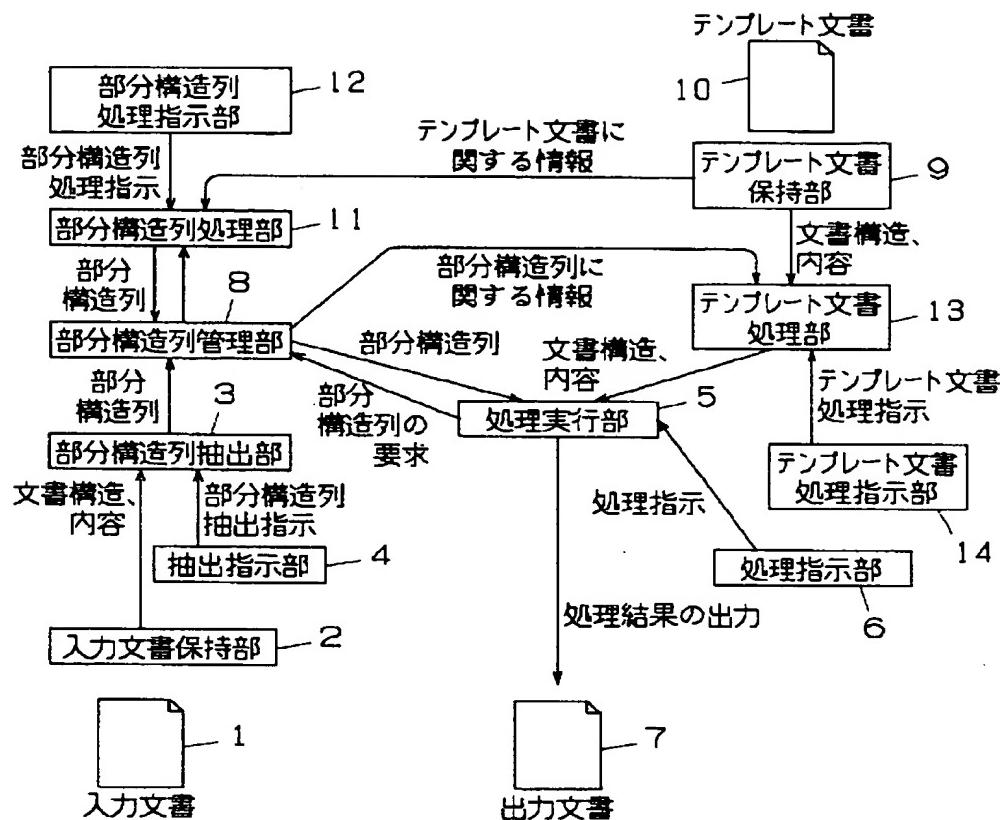
【図30】



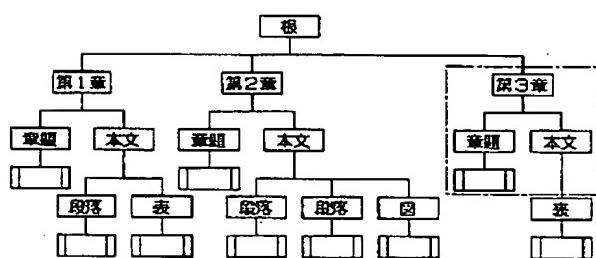
【図32】



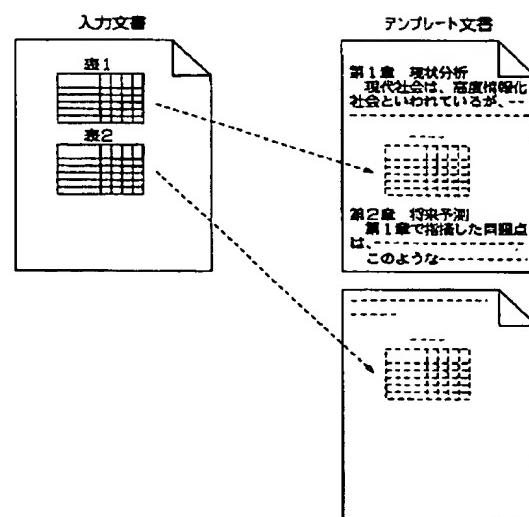
【図31】



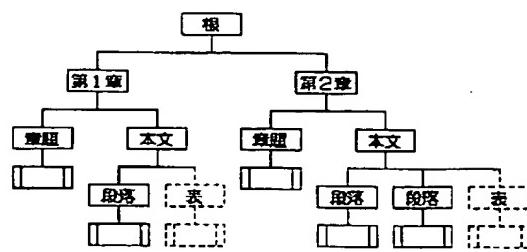
【図34】



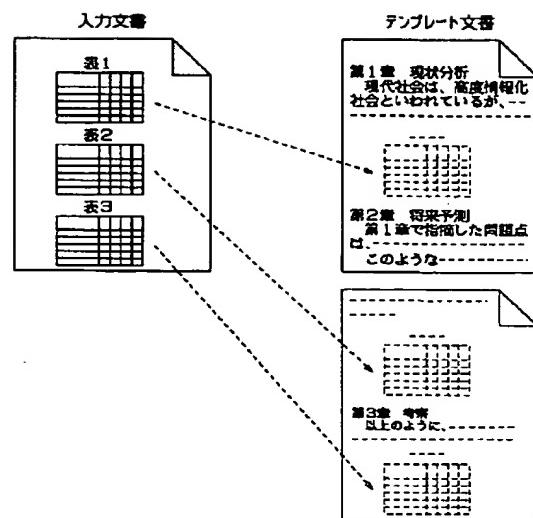
【図35】



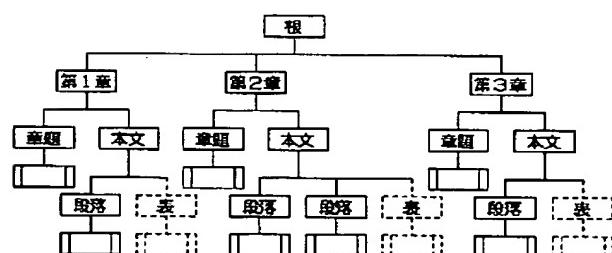
【図37】



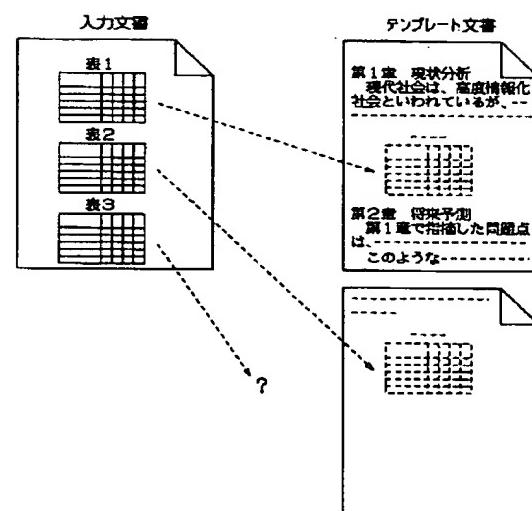
【図38】



【図40】



【図41】



【図42】

